DATA MINING WITH BIG DATA REVOLUTION HYBRID

R.Elankavi [1]* R.Kalaiprasath [1] R.Udayakumar [2]

[1]Research Scholar, Bharath University, Chennai, Asst. Professor, Aksheyaa college of Engineering, Chennai.

[2]Research Supervisor, Associate Professor, Department of Information Technology, Bharath University, Chennai.

Submitted: May 27, 2017 ***Accepted: June 15, 2017*** ***Published: Sep 1, 2017***

Abstract- Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This paper presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. We analyze the challenging issues in the data-driven model and also in the Big Data revolution.

Index terms: HACE, demand-driven, data storage

I. INTRODUCTION

The data held by Australian Government agencies is both a national and government asset. It is also a potential source of opportunity. In this context, Australian Government agencies, like many other organisations, are aware of the challenges and opportunities that big data represents to the way they develop policy and deliver services to citizens. The purpose of this issues paper is to provide an opportunity to consider the range of opportunities presented to agencies in relation to the use of big data, and the emerging tools that allow us to better appreciate what it tells us, in the context of the potential concerns that this might raise. As an example, one of the major challenges facing agencies here is to leverage the value of big data sets while ensuring they continue to protect the privacy rights of the Australian public. The Australian Government is committed to protecting citizen's rights to privacy, and as part of that commitment, has recently strengthened the provisions of the Privacy Act. The Australian Government Information Management Office (AGIMO) acknowledges that big data, and its associated analytical tools, can provide a challenge to these rights, but believe that, with proper considerations, agencies will be able to use big data to develop better policies and deliver better services without compromising the privacy rights of the public. Our aim is to ensure that the use of the new technology and tools supporting big data will deliver benefits while maintaining compliance with privacy. To this end AGIMO will be working closely with the Office of the Australian Information Commissioner (OAIC), the Attorney General's Department (AGD) and experts across the public and private sectors as it develops a big data strategy.

Where are we now?

Data is being produced at an ever increasing rate. This growth in data production is being driven by: individuals and their increased use of media; organizations; the switch from analogue to digital technologies; and The proliferation of internet connected devices and systems. Government agencies hold or have access to an ever increasing wealth of data including spatial and location data, as well as data collected from and by citizens. Experience suggests that such data can be utilised in ways that have the potential to transform service design and delivery so that personalised and streamlined services, that accurately and specifically meet individual's needs, can be delivered to them in a timely manner. Private sector organisations such as Google, Twitter and Facebook hold enormous data stores on Australian citizens and people across the world, and offer access to these on commercial terms. While needing to carefully consider the

veracity of this data, it may be that agencies could consider using this data as part of big data analytics projects. The ethical, privacy and security implications of decisions such as these will need to be carefully considered.

Why a big data strategy?

The development of a big data strategy was initiated by the APS ICT Strategy 2012 – 2015i (ICT Strategy) which highlighted the need for a strategy to enhance cross-agency data analytic capability for improved policy and service delivery.

As awareness of the benefits of big data increases there is likely to be an increase in public debate regarding the balance of benefits versus the challenges associated with the technology. Government agencies need to be in a position to consider external expert opinion and enunciate their own position on the use of the technology. Opportunities for Australian Government agencies The opportunity that big data presents to government agencies is in the potential to unlock the value and insight contained in the data agencies already hold via the transformation of information, facts, relationships and indicators. The value of big data for agencies is limited by their ability to effectively manage the volume, velocity and variety of big data and the ability to derive useful information from this data. With every opportunity there come challenges or barriers and agencies must overcome these to enable the benefits of big data to be realised.

Consideration of advances in big data technology has shown that it has potential to enhance the government's analysis capability in areas such citizen-centric service delivery. It is evident that big data also provides insights into social networks and relationships as well as allowing for the development of predictive models for a number of applications.

Of interest more broadly to agencies, big data analysis may provide profound insights into a number of key areas of society including health care, medical and other sciences, transport and infrastructure, education, communication, meteorology and social sciences.

What the future looks like

A successful big data strategy is expected to assist in realising each of the priority areas observed in the ICT Strategy. The delivery of better services Improved efficiency of government operations Open engagement Challenges Meeting the challenges presented by big data will be difficult. The volume of data is already enormous and increasing every day. The velocity of its generation and growth is increasing, driven in part by the proliferation of internet connected devices. Furthermore, the variety of data being generated is also expanding, and organisation's

capability to capture and process this data is limited. Current technology, architecture, management and analysis approaches are unable to cope with the flood of data, and organisations will need to change the way they think about, plan, govern, manage, process and report on data to realise the potential of big data.

Privacy, security and trust The Australian Government is committed to protecting the privacy rights of its citizens and has recently strengthened the Privacy Act (through the passing of the Privacy Amendment (Enhancing Privacy Protection) Bill 2012) to enhance the protection of and set clearer boundaries for usage of personal information.

The public trust in government agencies and systems needs to be maintained. As the volume of government data holdings increase, the trust that Australians have in these agencies and their ability to securely hold information of a personal nature can easily be affected by leakage of data or information into the public domain. Agencies need to be able to maintain the public's trust and will need to consider this issue at the forefront when developing secure systems for managing big data stores. Liaison with industry experts is an important first step in this process.

II. Data management and sharing

Accessible information is the lifeblood of a robust democracy and a productive economy.[ii] Government agencies realise that for data to have any value it needs to be discoverable, accessible and usable, and the significance of these requirements only increases as the discussion turns towards big data.

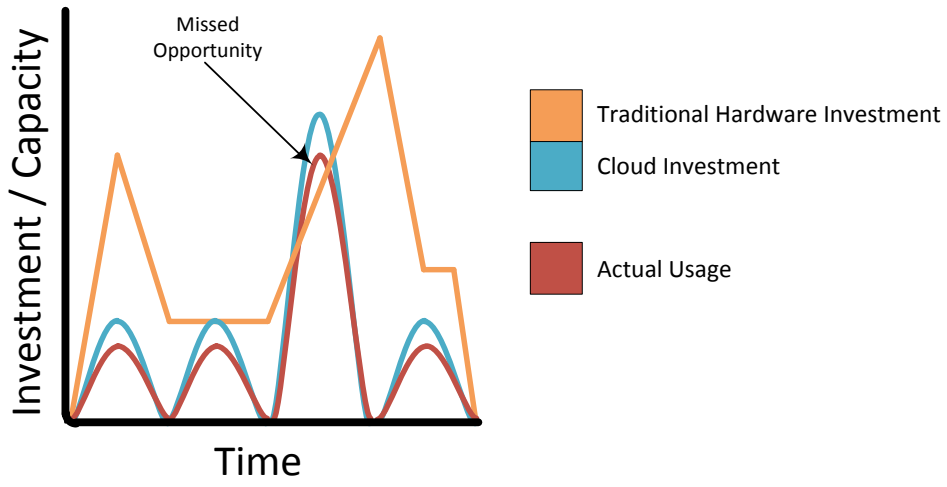
AGIMO is seeking the input and advice of the OAIC and other key agencies such as the AGD and the Defence Signals Directorate (DSD) in regards to big data and privacy for inclusion in the strategy.

III. Technology and analytical systems

The emergence of big data and the potential to undertake complex analysis of very large data sets is, essentially, a consequence of recent advances in the technology that allow this. If big data analytics is to be adopted by agencies, a large amount of stress may be placed upon current ICT systems and solutions which presently carry the burden of processing, analysing and archiving data. Government agencies will need to manage these new requirements efficiently in order to deliver net benefits through the adoption of new technologies.

Skills: Due to its relative youth and complexity, big data will require agencies to attract employees with diverse new skill sets. These skills include science, technological, research,

statistical, analytical and interpretive skills, business acumen and creativity — as well as an understanding of the underlying nature of the business process or policy intent. These skill sets are unlikely to be found in any one person, and this means that collaborative teams of specialists will need to be assembled to allow agencies to achieve optimal results from their data analysis efforts.



IV. DATA MINING CHALLENGES WITH BIG DATA:

For an intelligent learning database system [52] to handle Big Data, the essential key is to scale up to the exceptionally large volume of data and provide treatments for the characteristics featured by the aforementioned HACE theorem. Fig. 2 shows a conceptual view of the Big Data processing framework, which includes three tiers from inside out with considerations on data accessing and computing (Tier I), data privacy and domain knowledge (Tier II), and Big Data mining algorithms (Tier III). The challenges at Tier I focus on data accessing and arithmetic computing procedures. Because Big Data are often stored at different locations and data volumes may continuously grow, an effective computing platform will have to take distributed large-scale data storage into consideration for computing. For example, typical data mining algorithms require all data to be loaded into the main memory, this, however, is becoming a clear technical barrier for Big Data because moving data across different locations is expensive (e.g., subject to intensive network communication and other IO costs), even if we do have a super large main memory to hold all data for computing.

The challenges at Tier II center on semantics and domain knowledge for different Big Data applications. Such information can provide additional benefits to the mining process, as well as add technical barriers to the Big Data access (Tier I) and mining algorithms (Tier III). For example, depending on different domain applications, the data privacy and information sharing mechanisms between data producers and data consumers can be significantly different. Sharing sensor network data for applications like water quality monitoring may not be discouraged, whereas releasing and sharing mobile users' location information is clearly not acceptable for majority, if not all, applications. In addition to the above privacy issues, the application domains can also provide additional information to benefit or guide Big Data mining algorithm designs. For example, in market basket transactions data, each transaction is considered independent and the discovered knowledge is typically represented by finding highly correlated items, possibly with respect to different temporal and/or spatial restrictions. In a social network, on the other hand, users are linked and share dependency structures.

Big Privacy: Protecting Confidentiality in Big Data

Both the computer science and statistical science communities have developed a variety of criteria and methods for quantifying confidentiality risks. Indeed, a major thrust of research funded by the US National Science Foundation (including grants to us) is to integrate these two perspectives, taking the best of what both have to offer. In reviewing some of the risk metrics, we do not attempt to cover all approaches. Rather, we cover a few important ones that we are most familiar with. In statistical science, measures used in practice tend to be informal and heuristic in nature. For example, a common risk heuristic for publishing tabular magnitude data for business establishments (e.g., tables of total payroll within employee size groupings) is that no one establishment should contribute in excess of p% of the cell total, and no cell should comprise less than 3 establishments. Cells that do not meet these criteria are either suppressed or perturbed. The most general and mathematically formal method of disclosure risk assessment is based on Bayesian probabilities of re-identification, by which we means posterior probabilities that intruders could learn information about data subjects given the released data and a set of assumptions about the intruder's knowledge and behavior.

Agencies can compute these measures across a variety of intruder knowledge scenarios as a way of identifying particularly risky records and making an informed decision about data release policy in the face of uncertainty (the goal of statistical science in general). Computing these

probabilities in practice is computationally demanding and requires innovative methodology, especially for big data. In computer science, some of the early efforts to quantify confidentiality risk were targeted to thwart re-identification attacks (which we described in the introduction) by ensuring that no individual's record is unique in the data. This motivated a popular notion of privacy called K-Anonymity, which required that microdata be released in a manner that no individual's record is distinguishable from at least K-1 other records. While this seemingly avoids the privacy breaches discussed in the introduction, it has two drawbacks. An adversary (especially one with prior knowledge) can learn sensitive information. For instance, suppose a hospital releases K-anonymous microdata about patients, and you know your neighbor Bob is in the data. If individuals in the anonymous group containing Bob all have either cancer or the flu, and you know for a fact that Bob does not have the flu, then you can deduce that Bob has cancer. K-Anonymity has been extended in a number of ways to handle this shortcoming.

One example is L-Diversity, which requires that each group of individuals who are indistinguishable via quasi-identifiers (like age, gender, zip code, etc.) not share the same value for the sensitive attribute (like disease), but rather has L distinct well represented (of roughly same proportion) values. The current state of the art disclosure metric is called differential privacy. It eliminates (to a large extent) the confidentiality issues in K-anonymity, L-diversity and their extensions. Differential privacy can be best explained using the following opt-in/opt-out analogy. Suppose an agency (e.g., the Census Bureau or a search engine) wants to release microdata. Any individual has two options: opt-out of the microdata so that their privacy is protected, or opt-in and hope that an informed attacker can't infer sensitive information using the released microdata. A mechanism for microdata release is said to guarantee ε-differential privacy if for every pair of inputs D1 and D2 that differ in one individual's record (e.g., D1 contains record t and D2 does not contain t), and every microdata release M, the probability that the mechanism outputs M with input D1 should be close to (within an exp(ε) factor of) the probability that the mechanism outputs M with input D2. In this way, the release mechanism is insensitive to a single individual's presence (opt-in) or absence (opt-out) in the data. Thus, differential privacy represents a strong guarantee. Moreover, differential privacy satisfies an important property called composability -- if M1 and M2 are two mechanisms that satisfy differential privacy with parameters ε1 and ε2, then releasing the outputs of M1 and M2 together also satisfies differential privacy with parameter ε1+ε2. Other known privacy conditions (e.g. k-

anonymity and l-diversity) do not satisfy composability, and hence, two privacy preserving releases using these definitions can result in a privacy breach.

Collective Mining of Bayesian Networks from Distributed Heterogeneous Data

A collective approach to learning a Bayesian network from dis- tributed heterogenous data. In this approach, we rst learn a local Bayesian network at each site using the local data. Then each site identi es the observations that are most likely to be evidence of coupling between local and non-local variables and transmits a subset of these observations to a central site. Another Bayesian network is learnt at the central site using the data transmitted from the local site. The local and central Bayesian networks are combined to obtain a collective Bayesian network, that models the entire data. Raw data is useful only when it is transformed into knowledge or useful in- formation.

V. Privacy Preserving Data Mining

In this paper we address the issue of privacy preserving data mining. Specifically, we consider a scenario in which two parties owning confidential databases wish to run a data mining algorithm on the union of their databases, without revealing any unnecessary information. Our work is motivated by the need to both protect privileged information and enable its use for research or other purposes. The above problem is a specific example of secure multi-party computation and as such, can be solved using known generic protocols. However, data mining algorithms are typically complex and, furthermore, the input usually consists of massive data sets. The generic protocols in such a case are of no practical use and therefore more efficient protocols are required. Data mining is a recently emerging field, connecting the three worlds of Databases, Artificial Intelligence and Statistics.

The information age has enabled many organizations to gather large volumes of data. However, the usefulness of this data is negligible if “meaningful information” or “knowledge” cannot be extracted from it. Data mining, otherwise known as knowledge discovery, attempts to answer this need. In contrast to standard statistical methods, data mining techniques search for interesting information without demanding a priori hypotheses. As a field, it has introduced new concepts and algorithms such as association rule learning. It has also applied known machine-learning algorithms such as inductive-rule learning (e.g., by decision trees) to the setting where very large databases are involved. Data mining techniques are used in business and research and are becoming more and more popular with time

VI. Proposed System

In proposed system to build a stream-based Big Data analytic framework for fast response and real-time decision making.

The key challenges and research issues include: - designing Big Data sampling mechanisms to reduce Big Data volumes to a manageable size for processing; - building prediction models from Big Data streams.

Such models can adaptively adjust to the dynamic changing of the data.

A knowledge indexing framework to ensure real-time data monitoring and classification for Big Data applications.

Advantages

Hug data store and retrieve

Adapted all environments

More reliable

User friendly

Avoid collusions (eg. Dead lock)

Ignore network traffics.

Module Description

Distributed and Decentralized Control:

To share the information and multisystem and centralized database environment .

It provide and control the multiple process and store ,retrieve .

Complex and Evolving Relationships:

To analysis and avoid deadlock and optimized to the complex query to the user and provide multiple service to the user.

And it provide the relationship between the multiuser and multiple server throughout the network .

VII. Huge Data with Heterogeneous:

Anonymity data to store and indexing with the database and provide the service the user requirements.

Various type information to store and retrieve from the server by the help of big data mining.

Big data mining analysis:

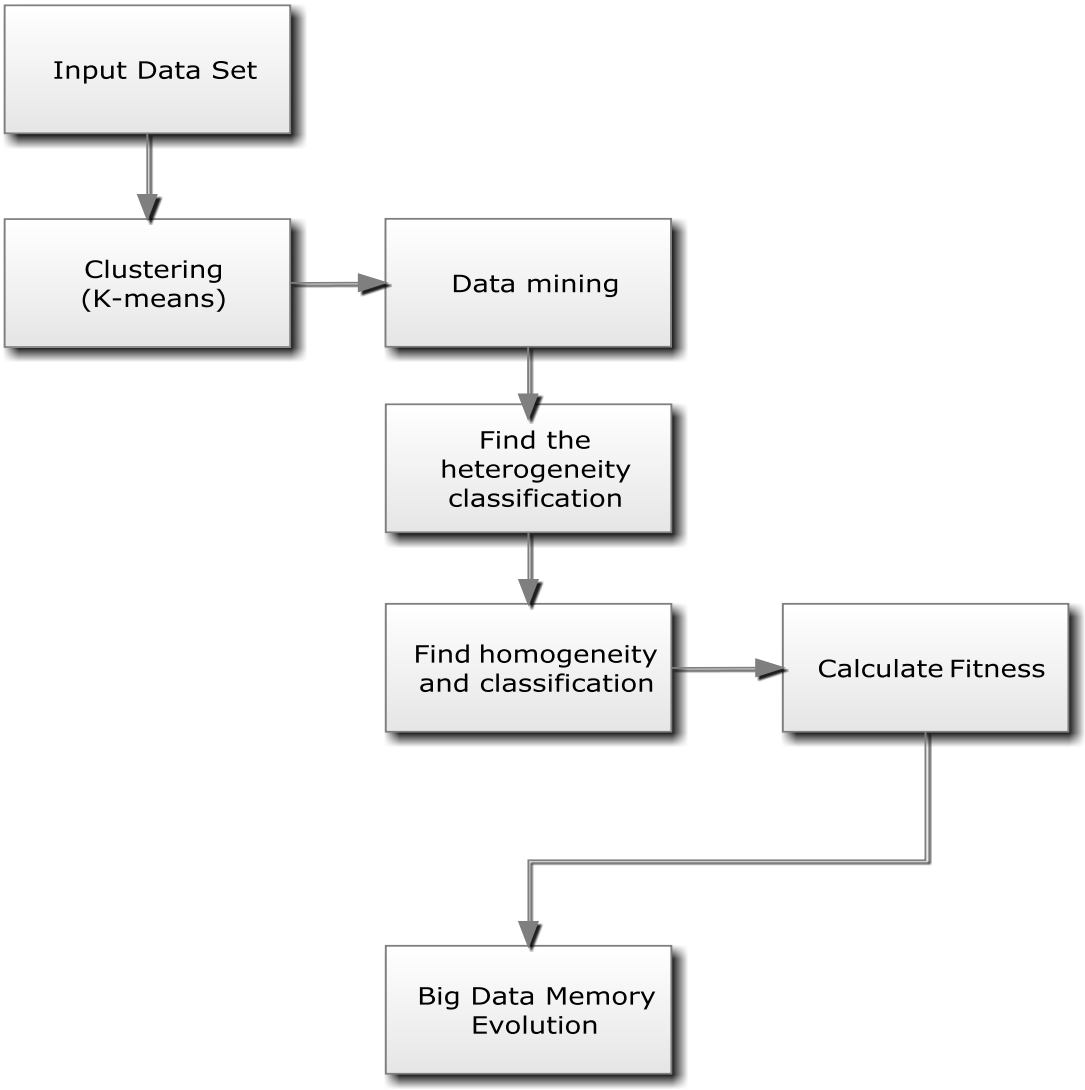
To clustering the data or information through out the one client to another client .

Extract the knowledge form the database by the help of mining.

Performance analysis:

Performance analysis to provide the information to the compare statement for the user by the help of diagrams.

Architecture diagram:



VIII. CONCLUSIONS:

Driven by real-world applications and key industrial stakeholders and initialized by national funding agencies, managing and mining Big Data have shown to be a challenging yet very compelling task. While the term Big Data literally concerns about data volumes, our HACE

theorem suggests that the key characteristics of the Big Data are 1) huge with heterogeneous and diverse data sources, 2) autonomous with distributed and decentralized control, and 3) complex and evolving in data and knowledge associations. Such combined characteristics suggest that Big Data require a "big mind" to consolidate data for maximum values [27]. To explore Big Data, we have analyzed several challenges at the data, model, and system levels. To support Big Data mining, high-performance computing platforms are required, which impose systematic designs to unleash the full power of the Big Data. At the data level, the autonomous information sources and the variety of the data collection environments, often result in data with complicated conditions, such as missing/uncertain values.

In other situations, privacy concerns, noise, and errors can be introduced into the data, to produce altered data copies. Developing a safe and sound information sharing protocol is a major challenge. At the model level, the key challenge is to generate global models by combining locally discovered patterns to form a unifying view. This requires carefully designed algorithms to analyze model correlations between distributed sites, and fuse decisions from multiple sources to gain a best model out of the Big Data. At the system level, the essential challenge is that a Big Data mining framework needs to consider complex relationships between samples, models, and data sources, along with their evolving changes with time and other possible factors. A system needs to be carefully designed so that unstructured data can be linked through their complex relationships to form useful patterns, and the growth of data volumes and item relationships should help form legitimate patterns to predict the trend and future. We regard Big Data as an emerging trend and the need for Big Data mining is arising in all science and engineering domains. With Big Data technologies, we will hopefully be able to provide most relevant and most accurate social sensing feedback to better understand our society at realtime.

REFERENCES

[1] Aizat Azmi, Ahmad Amsyar Azman, Sallehuddin Ibrahim, and Mohd Amri Md Yunus, "Techniques In Advancing The Capabilities Of Various Nitrate Detection Methods: A Review", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 2, June 2017, pp. 223-261.

[2] Tsugunosuke Sakai, Haruya Tamaki, Yosuke Ota, Ryohei Egusa, Shigenori Inagaki, Fusako Kusunoki, Masanori Sugimoto, Hiroshi Mizoguchi, "Eda-Based Estimation Of Visual Attention

By Observation Of Eye Blink Frequency", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 2, June 2017, pp. 296-307.

[3] Ismail Ben Abdallah, Yassine Bouteraa, and Chokri Rekik , "Design And Development Of 3d Printed Myoelctric Robotic Exoskeleton For Hand Rehabilitation", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 2, June 2017, pp. 341-366.

[4] S. H. Teay, C. Batunlu and A. Albarbar, "Smart Sensing System For Enhanceing The Reliability Of Power Electronic Devices Used In Wind Turbines", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 2, June 2017, pp. 407- 424

[5] SCihan Gercek, Djilali Kourtiche, Mustapha Nadi, Isabelle Magne, Pierre Schmitt, Martine Souques and Patrice Roth, "An In Vitro Cost-Effective Test Bench For Active Cardiac Implants, Reproducing Human Exposure To Electric Fields 50/60 Hz", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 1, March 2017, pp. 1- 17

[6] P. Visconti, P. Primiceri, R. de Fazio and A. Lay Ekuakille, "A Solar-Powered White Led-Based Uv-Vis Spectrophotometric System Managed By Pc For Air Pollution Detection In Faraway And Unfriendly Locations", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 1, March 2017, pp. 18- 49

[7] Samarendra Nath Sur, Rabindranath Bera and Bansibadan Maji, "Feedback Equalizer For Vehicular Channel", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 1, March 2017, pp. 50- 68

[8] Yen-Hong A. Chen, Kai-Jan Lin and Yu-Chu M. Li, "Assessment To Effectiveness Of The New Early Streamer Emission Lightning Protection System", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 1, March 2017, pp. 108- 123

[9] Iman Heidarpour Shahrezaei, Morteza Kazerooni and Mohsen Fallah, "A Total Quality Assessment Solution For Synthetic Aperture Radar Nlfm Waveform Generation And Evaluation In A Complex Random Media", International Journal on Smart Sensing and Intelligent Systems., VOL. 10, NO. 1, March 2017, pp. 174- 198

[10] P. Visconti ,R.Ferri, M.Pucciarelli and E.Venere, "Development And Characterization Of A Solar-Based Energy Harvesting And Power Management System For A Wsn Node Applied To Optimized Goods Transport And Storage", International Journal on Smart Sensing and Intelligent Systems., VOL. 9, NO. 4, December 2016 , pp. 1637- 1667

[11] YoumeiSong,Jianbo Li, Chenglong Li, Fushu Wang, “Social Popularity Based Routing In Delay Tolerant Networks”, International Journal on Smart Sensing and Intelligent Systems., VOL. 9, NO. 4, December 2016 , pp. 1687- 1709

[12] Seifeddine Ben Warrad and OlfaBoubaker, “Full Order Unknown Inputs Observer For Multiple Time-Delay Systems”, International Journal on Smart Sensing and Intelligent Systems., VOL. 9, NO. 4, December 2016 , pp. 1750- 1775

[13] Rajesh, M., and J. M. Gnanasekar. "Path observation-based physical routing protocol for wireless ad hoc networks." International Journal of Wireless and Mobile Computing 11.3 (2016): 244-257.

[14]. Rajesh, M., and J. M. Gnanasekar. "Congestion control in heterogeneous wireless ad hoc network using FRCC." Australian Journal of Basic and Applied Sciences 9.7 (2015): 698-702.

[15]. Rajesh, M., and J. M. Gnanasekar. "GCCover Heterogeneous Wireless Ad hoc Networks." Journal of Chemical and Pharmaceutical Sciences (2015): 195-200.

[16]. Rajesh, M., and J. M. Gnanasekar. "CONGESTION CONTROL USING AODV PROTOCOL SCHEME FOR WIRELESS AD-HOC NETWORK." Advances in Computer Science and Engineering 16.1/2 (2016): 19.

[17]. Rajesh, M., and J. M. Gnanasekar. "An optimized congestion control and error management system for OCCEM." International Journal of Advanced Research in IT and Engineering 4.4 (2015): 1-10.

[18]. Rajesh, M., and J. M. Gnanasekar. "Constructing Well-Organized Wireless Sensor Networks with Low-Level Identification." World Engineering & Applied Sciences Journal 7.1 (2016).

[19] Aftab Ali Haider, AcmerNadeem, ShamailaAkram, “Safe Regression Test Suite Optimization: A Review”,In: Proc. of IEEE International Conference on Open Source Systems and Technologies, pp. 7-12, 2016.

[20] AvinashGupta,Namita Mishra,Dharmender Singh Kushwaha, “Rule-Based test case Reduction Technique using Decision Table”,In: Proc. of IEEE Conference on International Advance Computing Conference,pp.1398-1405,2014.

[21] Annibalepanichella,Rocco oliveto,Massimiliano Di Penta,Andrea De Lucia, “ Improving multi-objective test case Selection by Injecting Diversity in genetic Algorithms”, IEEE Transactions on Software Engineering,pp.358-383,Vol.41,No.4,April 2015.

[22] Zhang Hui, “Fault Localization Method Generated by Regression Test Cases on the Basis of Genetic Immune Algorithm”, In: proc. Of IEEE conference on Annual International Computers, Software & Applications Conference, pp. 46-51, 2016.

[23] S. Yoo and M. Harman, “Regression testing minimization, selectionand prioritization: A survey,” Softw. Test. Verif. Rel., vol. 22,no. 2, pp. 67–120, Mar. 2012.

[24] S. Yoo, “A novel mask-coding representation for set cover problemswith applications in test suite minimisation,” In: Proc. of 2nd International Symposium. Search-Based Software. Eng., 2010, pp. 19–28.

[25] S. Yoo and M. Harman, “Pareto efficient multi-objective test case selection,” In: Proc. of ACM /SIGSOFT Int. Symp. Softw. Testing Anal.,2007, pp. 140–150.

[26] S. Yoo and M. Harman, “Using hybrid algorithm for Pareto efficientmulti-objective test suite minimisation,” J. Syst. Softw.,vol. 83, no. 4, pp. 689–701, 2010.

[27] S. Yoo, M. Harman, and S. Ur, “Highly scalable multi objectivetest suite minimization using graphics cards,” In:Proc. of 3rd Int.Conf. Search Based Softw. Eng., 2011, pp. 219–236.

[28] Q. Zhang and Y.-W. Leung, “An orthogonal genetic algorithm for multimedia multicast routing,” IEEE Trans. Evol. Comput., vol. 3,no. 1, pp. 53–62, Apr. 1999.

[29] J. Zhu, G. Dai, and L. Mo, “A cluster-based orthogonal multi objective genetic algorithm”, Comput. Intell. Intell. Syst., vol. 51,pp. 45–55, 2009.

[30] E. Zitzler, D. Brockhoff, and L. Thiele, “The hypervolume indicatorrevisited: On the design of Pareto-compliant indicators via weighted integration”,In: Proc. of 4th Int. Conf. Evol. Multi-CriterionOptim., 2007, pp. 862–876.

[31] Jones JA, Harrold MJ. “Empirical Evaluation of the Tarantula Automatic Fault - Localization Technique”. In: Proc. of 20th IEEE/ ACM International Conference on Automated Software Engineering, 2005: 273-282.

[32] Jones JA, Harrold MJ, Stasko J. “Visualization of Test Information to Assist Fault Localization”.In: Proc. ofthe 24th International Conference on Software Engineering, 2002:467-477.